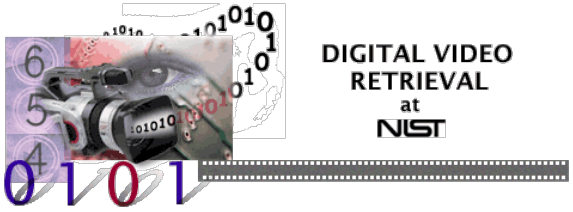
Multimedia Event Detection and Recounting Task Reports

Time		Presentation
9:30	– 9:50	University of Amsterdam (MediaMill)
9:50	– 10:10	TNO (TNO)
10:10	– 10:30	Access to Audiovisual Media (AXES)
10:30	– 11:00	Break in the NIST West Square Cafeteria
11:00	– 11:15	Carnegie Mellon University (CMU)
11:15	– 11:45	Raytheon (BBN-VISER)
11:45	– 12:15	Kitware, Inc. (GENIE)
12:15	– 12:35	SRI International, University of Amsterdam (SRI_SESAME)
12:35	– 1:40	**Lunch** is served in the NIST West Square Cafeteria (please have your lunch ticket ready)
1:40	– 2:10	SRI, Sarnoff, Central Fl. U., U. Mass., Cycorp, ICSI, Berkeley (SRIAURORA)
2:10	– 2:40	Discussion
2:40	– 3:00	Poster and demo boaster session (Alan Smeaton)
3:00	– 3:20	**Break** with refreshments served in the Green Auditorium area
3:20	– 4:30	Posters and demos in hallway by the Green Auditorium

2013 TRECVID Workshop Multimedia Event Detection and Multimedia Event Recounting Tasks

Jonathan Fiscus, Greg Sanders, **David Joy**, Paul Over
National Institute of Standards and Technology (NIST)

Martial Michel
Systems Plus, Inc.



Talk Outline

- HAVIC Data Resources
- MED Task Overview and Results
- MER Task Overview and Results

MED Task Definition

Given an event specified by an **event kit**, search multimedia recordings for the event:

1. Rank each clip in the collection,
2. Measure the search collection metadata generation time
3. Measure the event training metadata generation time
4. Measure the event detection execution time

An MED Event is

- complex activity occurring at a specific place and time;
- involves people interacting with other people and/or objects;
- consists of a number of human actions, processes, and activities that are loosely or tightly organized and that have significant temporal and semantic relationships to the overarching activity;
- is directly observable.

Rock Climbing Event Kit Text

Definition:

One or more people climb up or across rock formations or artificial rock walls.

Explication:

Rock climbing is a physically intense activity, where the goal is to reach the top or endpoint of a pre-defined route on a rock formation or artificial rock wall by finding a grip on the surface using hands and feet, and then pulling up using their arm and leg strength. …

Evidential Description:

- scene: outdoors in natural setting, indoors in rock climbing gym, or outdoors on a specially …
- objects/people: carabiners, rope, helmet, harness, rock formation, artificial rock wall, climbers
- activities: hooking rope to harness, moving hands and feet along side of rock face, grabbing rock …….
- audio: carabiners clinking, climbers making comments on the difficulty of the climb, onlookers cheering on …

Illustrative Examples

- Positive instances of the event
- Non-Positive “miss” clips that do not contain the event

Positive Rock Climbing Video Example

MED Evaluation Conditions

- MED Tasks
  - **Pre-Specified Event (PS)** – MED metadata generation optimized with knowledge of events
  - **Ad-Hoc Event (AH)** – MED metadata generation complete before events revealed
- Event Training Condition
  - **100Ex** – Use the event kit texts, 100 positives and 50 miss exemplars
  - **10Ex** – Use a 10 positive and 10 miss clip subset (20 total per event) of 100Ex
  - **0Ex** – Use Only the event kit texts
- System Submissions
  - One run per task/event training condition
    - **FullSys** – The teams system with all available technologies combined
  - Four prescribed contrast runs per task/event training condition
    - **OCRSys** – Optical Character Recognition only
    - **ASRSys** – Automatic Speech Recognition only
    - **VisualSys** – Non-OCR visual only
    - **AudioSys** – Non-ASR audio only
- Search video set
  - **PROGAll** -> The full Progress Collection set (98,000 clips, 3,722 hours)
  - **PROGSub** -> A 1/3 subset of PROGFull (32,000 clips, 1,243 hours)

The TRECVID MED 2013 Events

Pre-Specified Events

MED '11 Events

Changing a vehicle tire
Getting a vehicle unstuck
Grooming an animal
Making a sandwich
Parkour
Repairing an appliance
Working on a sewing project
Birthday party
Flash mob gathering
Parade

MED '12 Events

Attempting a bike trick
Cleaning an appliance
Dog show
Giving directions to a location
Marriage proposal
Renovating a home
Rock climbing
Town hall meeting
Winning a race without a vehicle
Working on a metal crafts project

Ad Hoc Events

New Events

Beekeeping
Wedding shower
Non-motorized veh. repair
Fixing musical instrument
Horse riding competition
Felling a tree
Parking a vehicle
Playing fetch
Tailgating
Tuning musical instrument

18 MED 2013 Finishers and Number of Runs

	Team	Ad-Hoc			Pre-Specified			Organization	MER Participation	
		100Ex	10Ex	0Ex	100Ex	10Ex	0Ex		2012	2013
3 Years	BBNVISER	5	5	5	5	5	5	Raytheon BBN Technologies, UMD, Columbia, UCF Team	✓	✓
	CERTH-ITI	2	2		2	2		Informatics and Telematics Inst., Centre for Research and Tech.	✓	✓
	CMU	5	5	5	5	5	5	Carnegie Mellon University	✓	✓
	Genie	4	4	4	4	4	4	Kitware Inc.	✓	✓
	IBM-Columbia	3	3	3	3	3	3	IBM T. J. Watson Research Center	✓	✓
	MediaMill	3	3		3	3		University of Amsterdam	✓	
	NII	3			3			National Institute of Informatics		
	Sesame	5	5	4	5	5	4	SRI International SESAME	✓	✓
	SRIAURORA	5	5	5	5	5	5	SRI International Sarnoff Aurora	✓	✓
	TokyoTechCanon	1			1			Tokyo Institute of Technology and Canon		
2 Years	AXES	5			1			Dublin City Univ, Univ Twente, Univ of Oxford, INRIA, Fraunhofer IAIS, Katholieke Universitiet Leuven, Technicolor, Erasmus University Rotterdam, Cassidian, BBC, Deutsche Welle, Beeld et. Geluid (NISV), ERCIM	✓	✓
	VIREO				2			City University of Hong Kong	✓	✓
1 Year	ORAND	2			1			ORAND		
	PicSOM	3	3		2	2		Aalto University		✓
	SiegenKobe Muro				2			Institute for Vision and Graphics, University of Siegen Graduate School of System Informatics, Kobe University College of Information and Systems, Muroran Institute of Technology		
	TNO	2		1	2		1	TNO		
	UMass	1		4	1		4	University of Mass.		
	VisQMUL	3	3	3	1	1	1	Queen Mary, University of London		
		52	**38**	**34**	**48**	**35**	**32**		**10**	**10**

HAVIC Data Resources

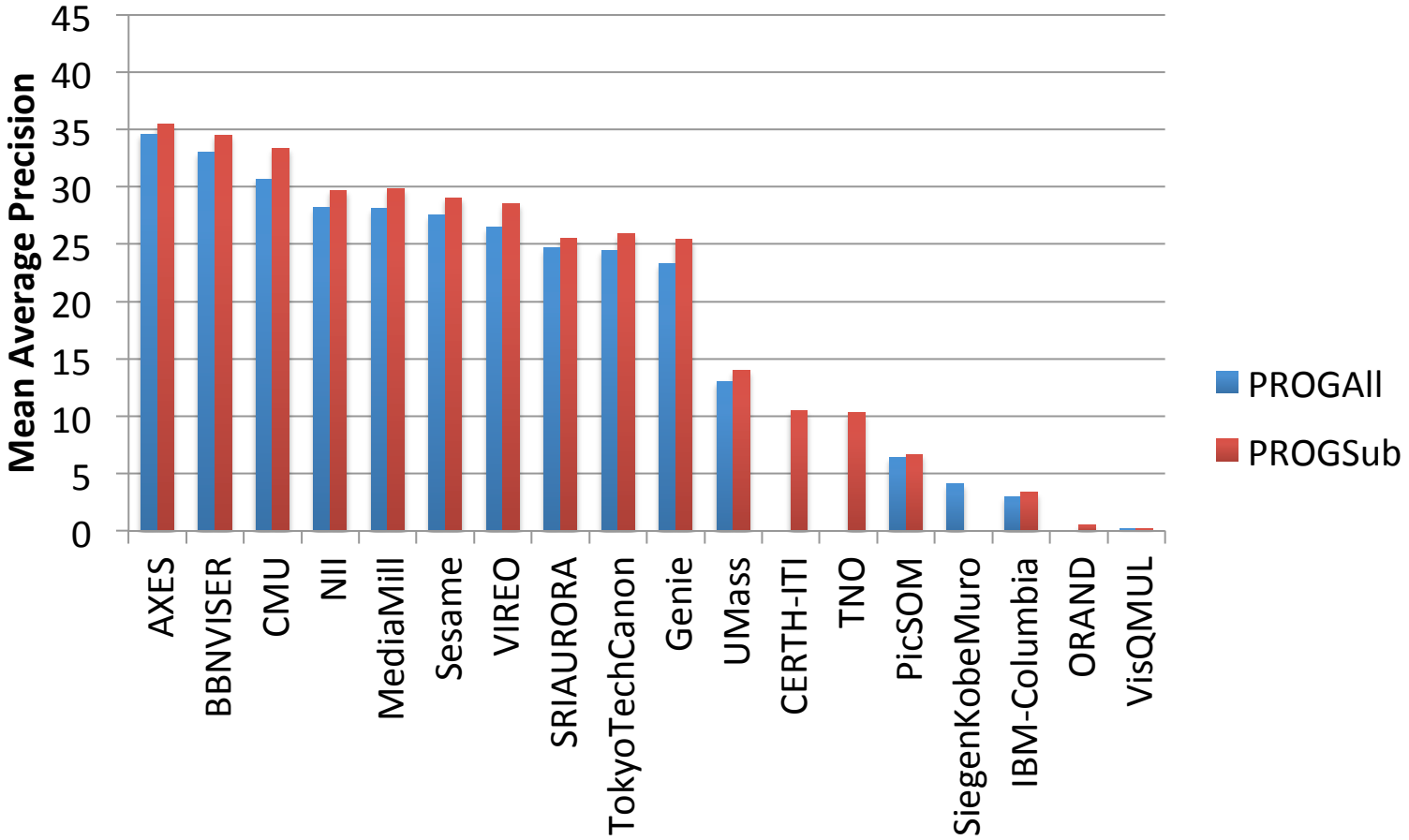
		Video clips	Video duration
Development Data	RESEARCH	10,000	314 hours
	10 Event Kits	1,400	74 hours
	Transcription	1,500	45 hours
Event Training Data	Event Background	5,000	146 hours
	30 Event Kits	3,000	180 hours
Test Data	MEDTest	27,000	838 hours
	KindredTest	14,500	675 hours
Evaluation Data	PROGAll	98,000	3,722 hours
	PROGSub	33,000	1,244 hours
	Total	144,049	5,840 hours

MED '13 Results

- New primary metrics
  - Mean Average Precision
  - (diagnostic) Minimal Acceptable Recall – R_0
- Computation speed and resource measurements
  - Metadata generation processing speed
  - Event query generation processing speed
  - Event search processing speed

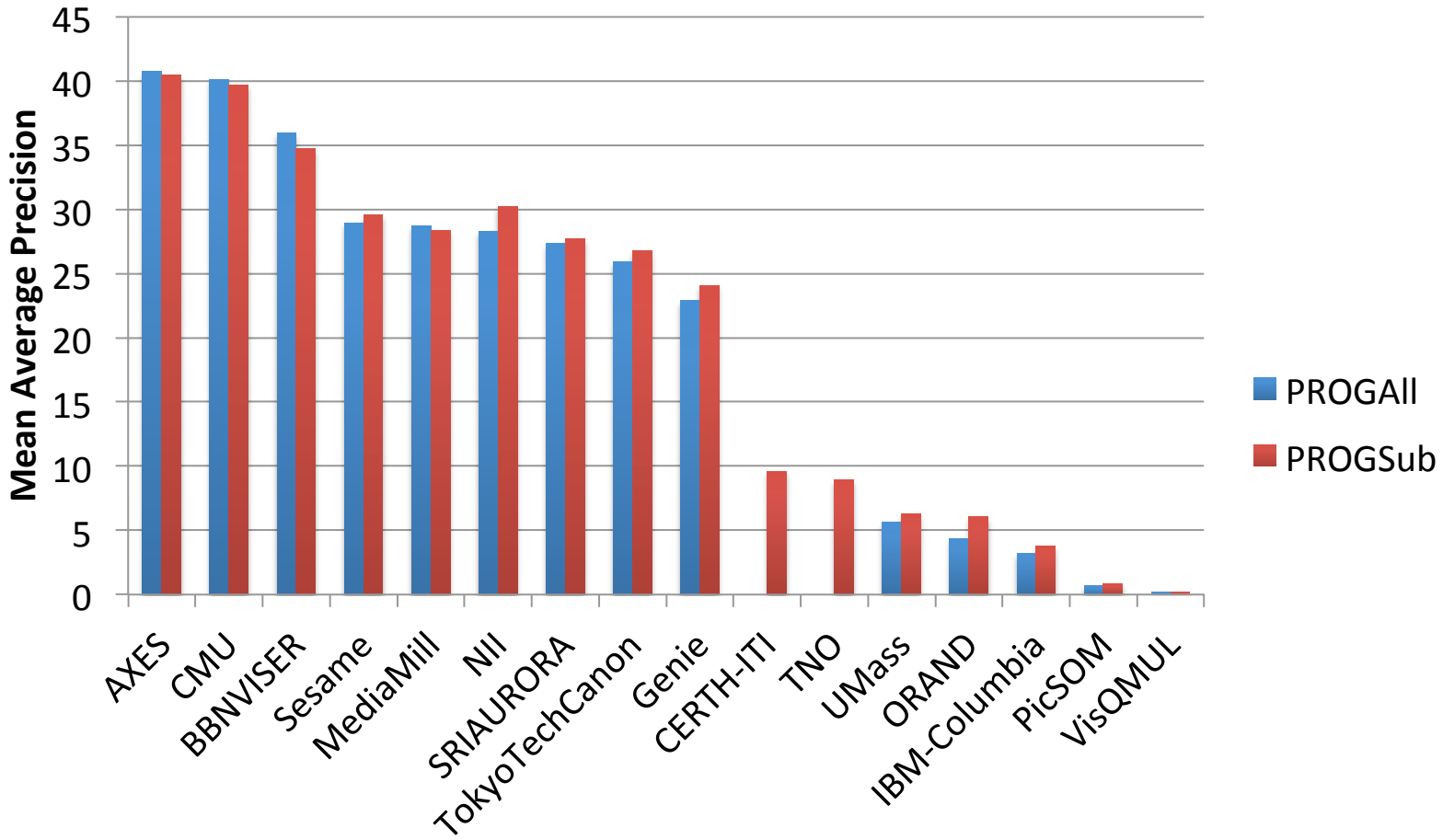
MED '13 Pre-Specified Event Task

Events E0[06-15,21-30], 100Ex, FullSys, All Teams



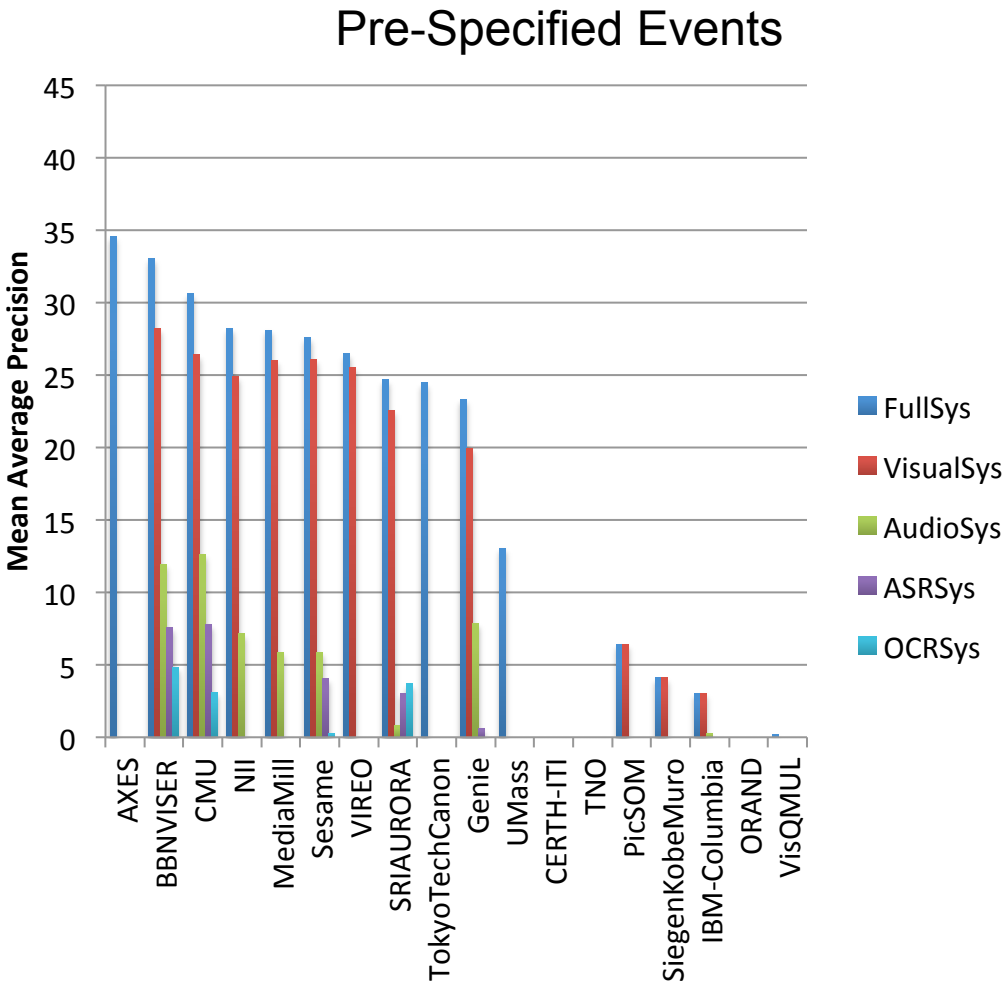
Post-Adjudicated MED '13 Ad-Hoc Event Task

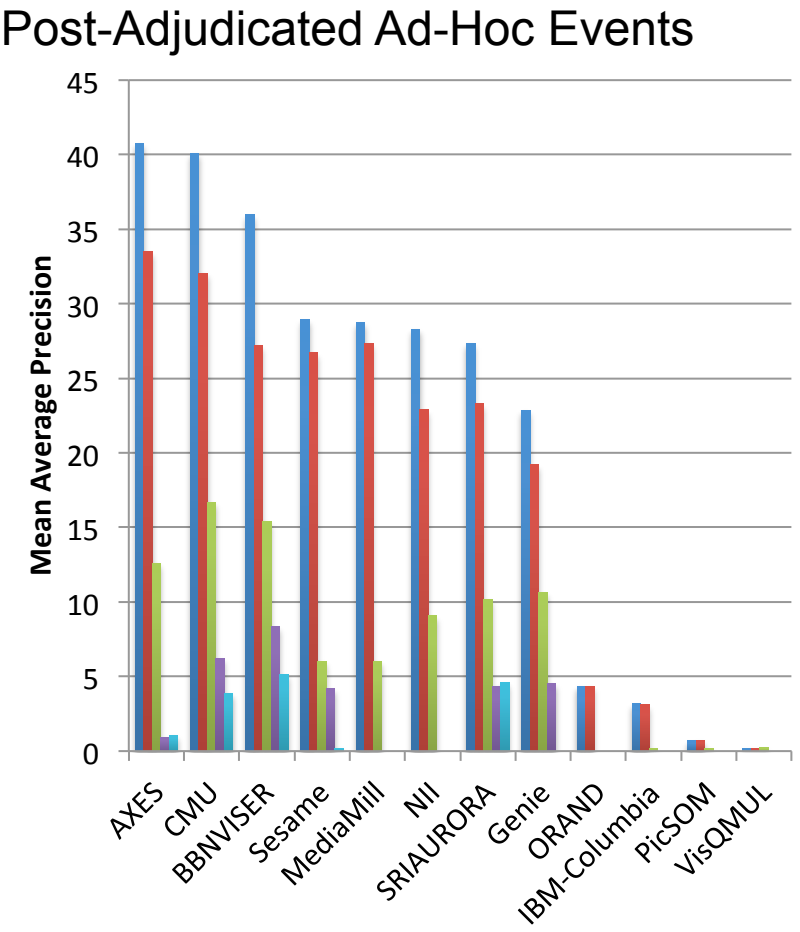
Events E031-E040, 100Ex, FullSys, All Teams



Effect of System Type

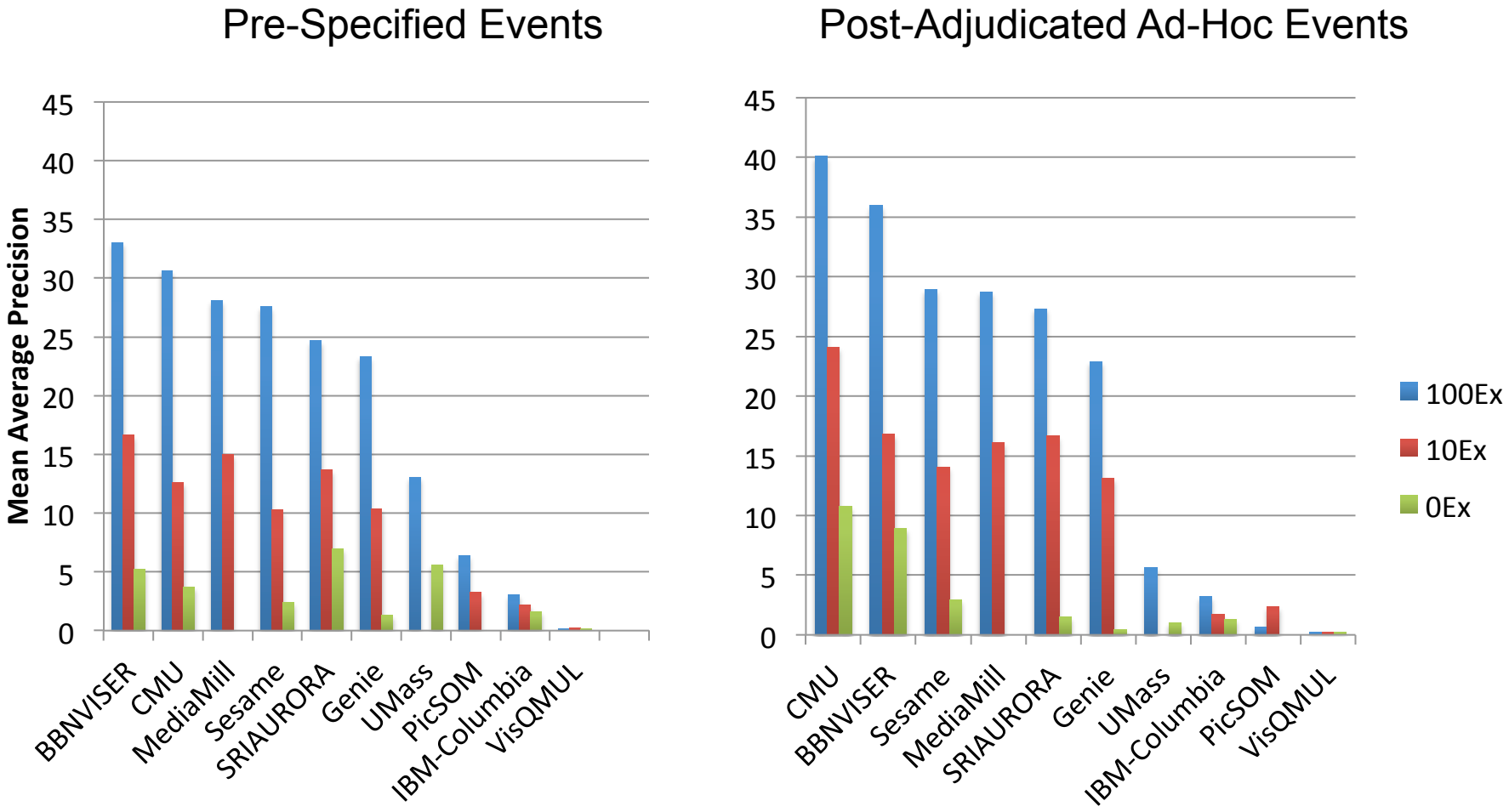
FullSys, VisualSys, AudioSys, ASRSys, OCRSys



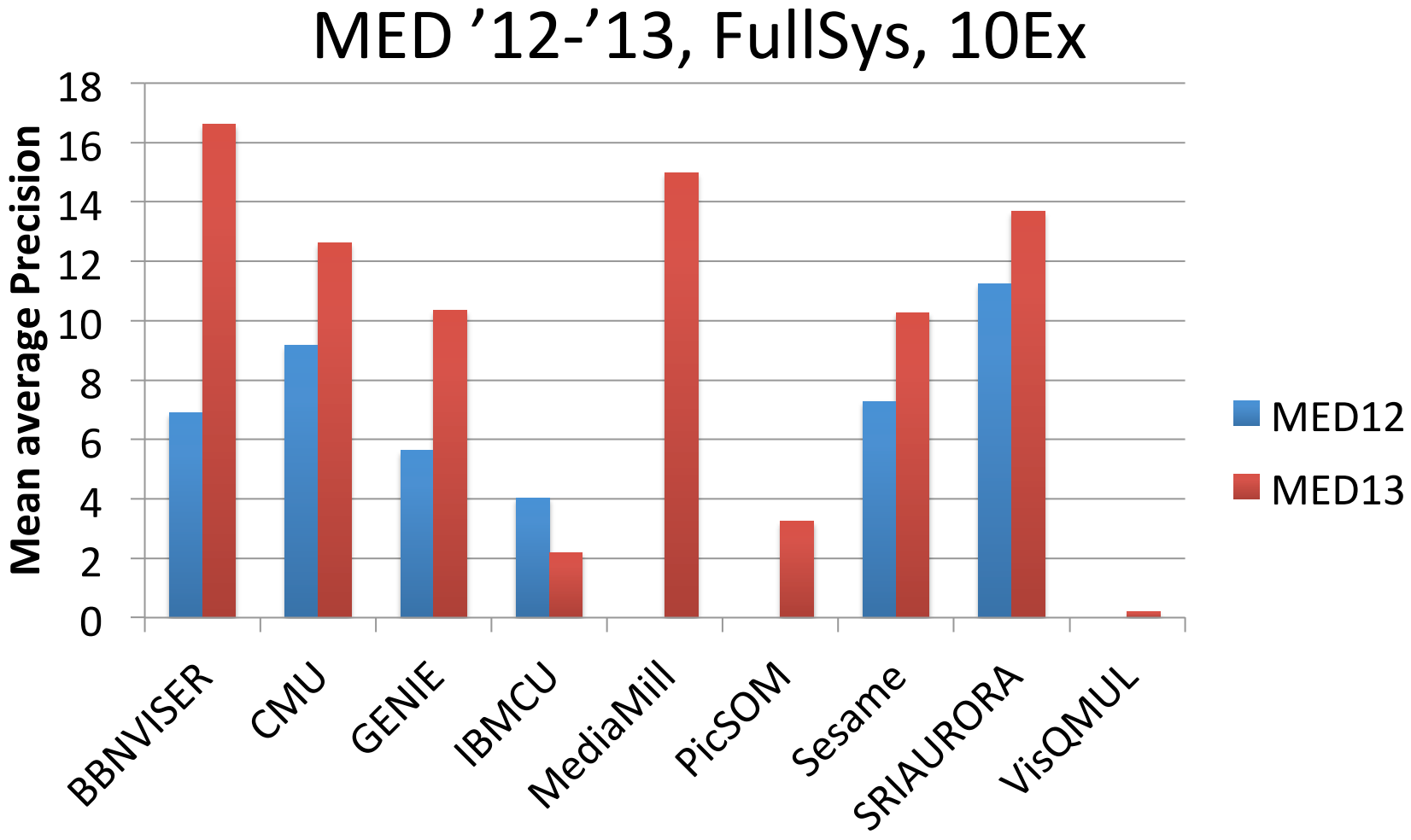


Effect of Training Exemplars

FullSys vs. 100EX, 10Ex, 0Ex

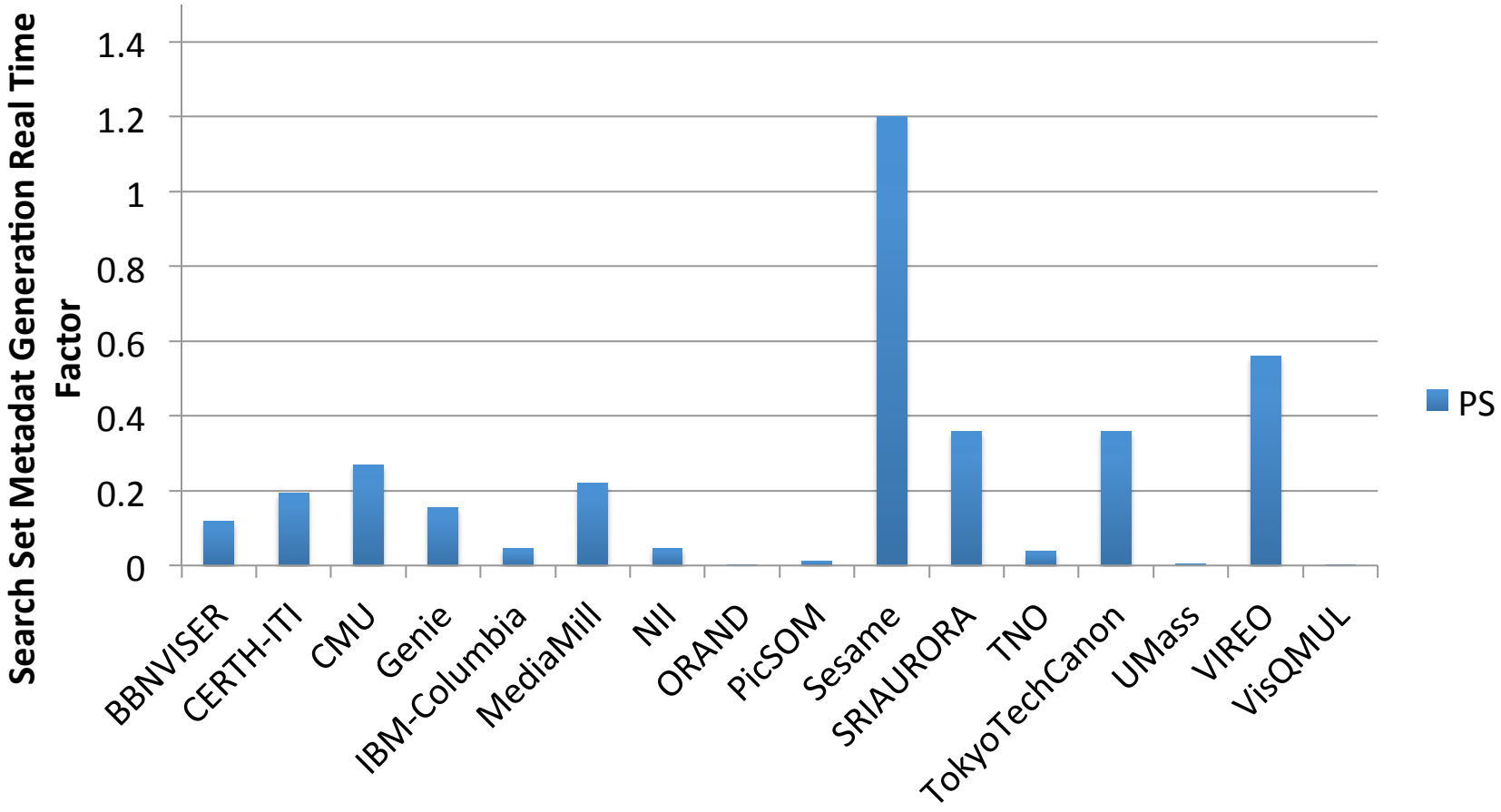


Historical MED MAP Scores



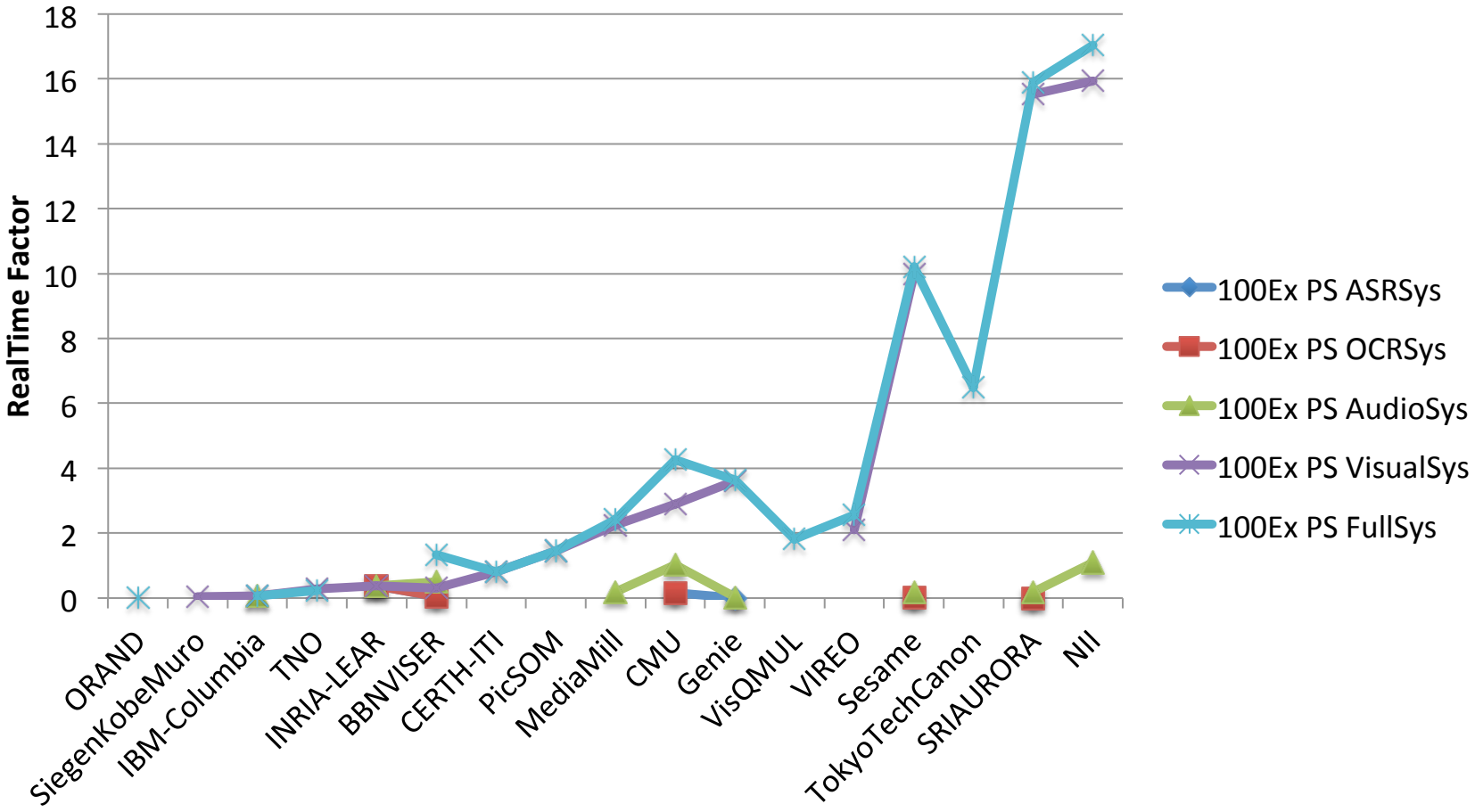
Operation Speeds:

Search Metadata Generation Real Time Factor



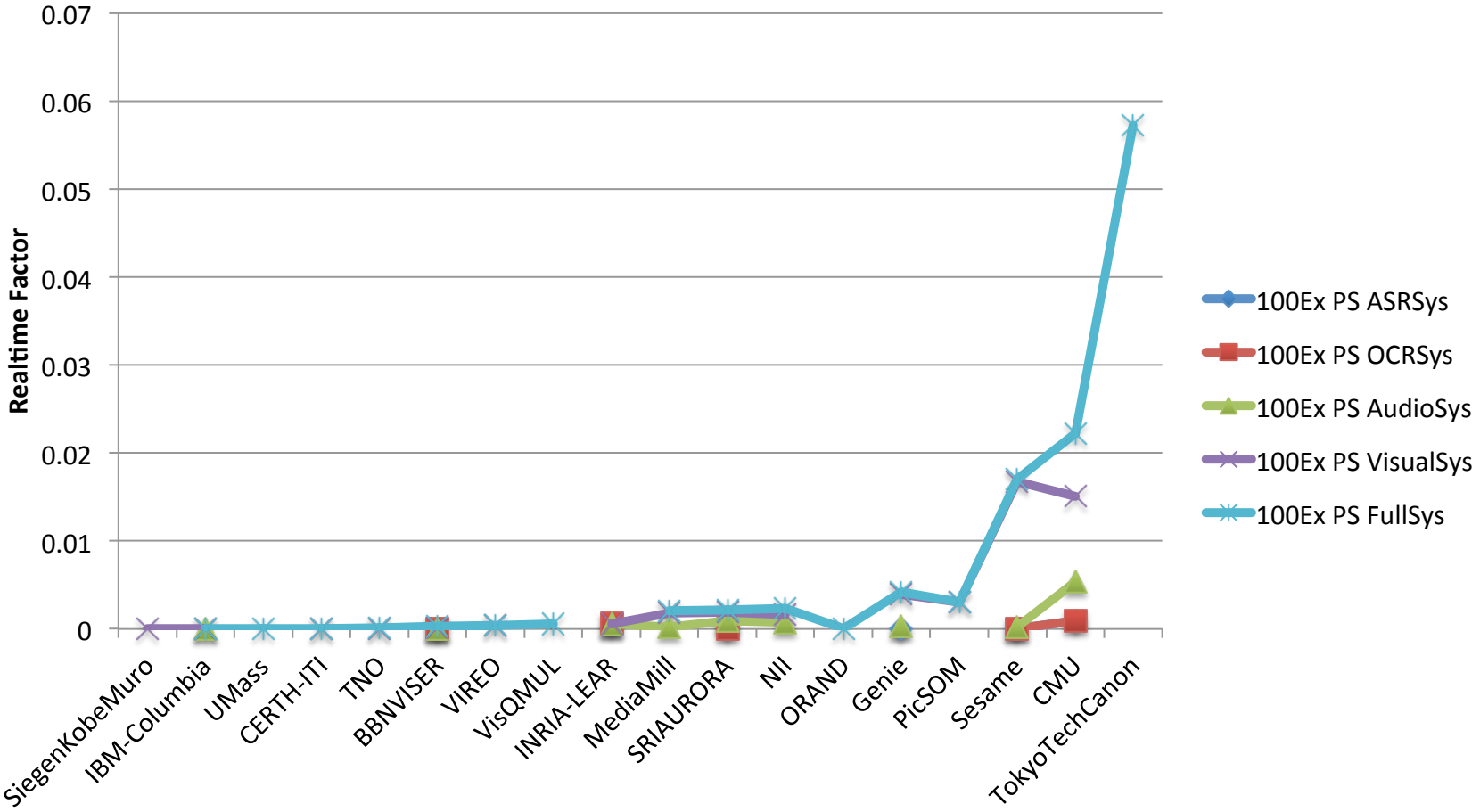
Operation Speeds: PS, 100 Exemplar

Event Agent Generation Real Time Factor



Operation Speeds: PS, 100 Exemplar

Detection Real Time Factor



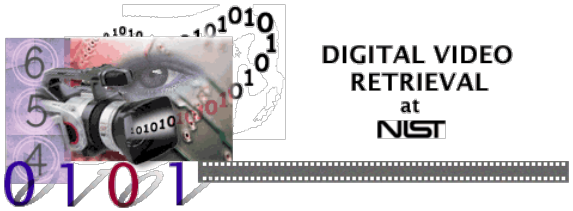
Summary/Conclusions

- Successful full evaluation of AdHoc events
  - AdHoc performance is higher than Pre-Specified in general
  - Switch to MAP has gone smoothly
- Questions to consider:
  - Are the the 100Ex, 10Ex, and 0Ex training exemplar conditions testing the parameter space adequately?
  - Are the contrast conditions useful?
  - Is the size of the test corpus too large?
  - Can we improve timing measures?
  - Should we make 10Ex the required condition for both PS and AdHoc?
  - Can we run a timed ad-hoc condition?
    - Checkout the event kit, build event model, submit result
- Known changes for MED 2014
  - A new test collect will be released – comparable in size to PROGAll
  - MED 13 AdHoc events will become PreSpecified Events next year

Questions?

Multimedia Event Recounting (MER) Task Overview and Results

David Joy, Greg Sanders, Jonathan Fiscus
NIST



The MER Task

- Recount the multimedia evidence that led the Multimedia Event Detection (MED) system to the decision that a particular multimedia clip contains an instance of a specific event.
- Recountings generated for the following MED evaluation condition
  - MED Task - Pre-Specified Event (PS)
  - Event Training Condition – 100Ex
  - PROGAll or PROGSub
  - System Submission – FullSys

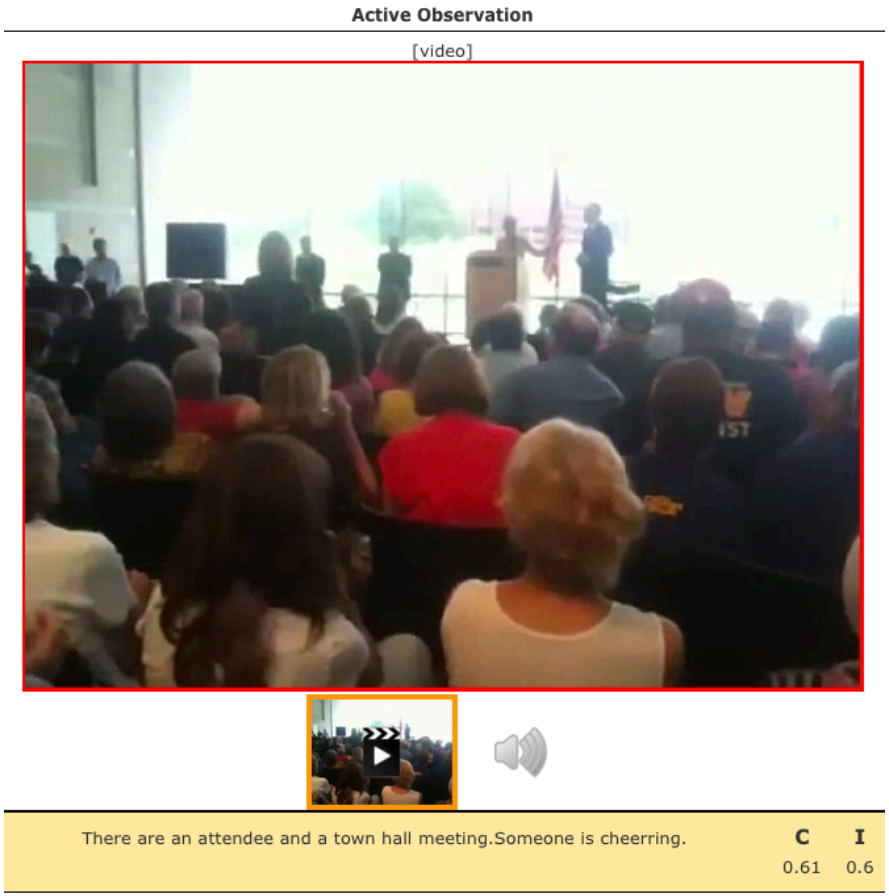
The Events

- An event is defined by an Event Kit
- Total of ten events chosen for this year
  - The five evaluated in 2012..
    - E022 Cleaning an appliance
    - E026 Renovating a home
    - E027 Rock climbing
    - E028 Town hall meeting
    - E030 Working on a metal crafts project
  - Five additional chosen..
    - E007 Changing a vehicle tire
    - E009 Getting a vehicle unstuck
    - E010 Grooming an animal
    - E013 Parkour
    - E015 Working on a sewing project

Recounting Structure

System output at a glance

- Recounting
  - EventID
  - ClipID
  - Observations (1 or more)
    - Description (text) – The "what"
    - Confidence
    - Importance
    - Snippets (1 or more) (Audio/Video/ Keyframe)
      - Begin & End Time – The "when"
      - Bounding Box (Video/ Keyframe) – The "where"



Judgment

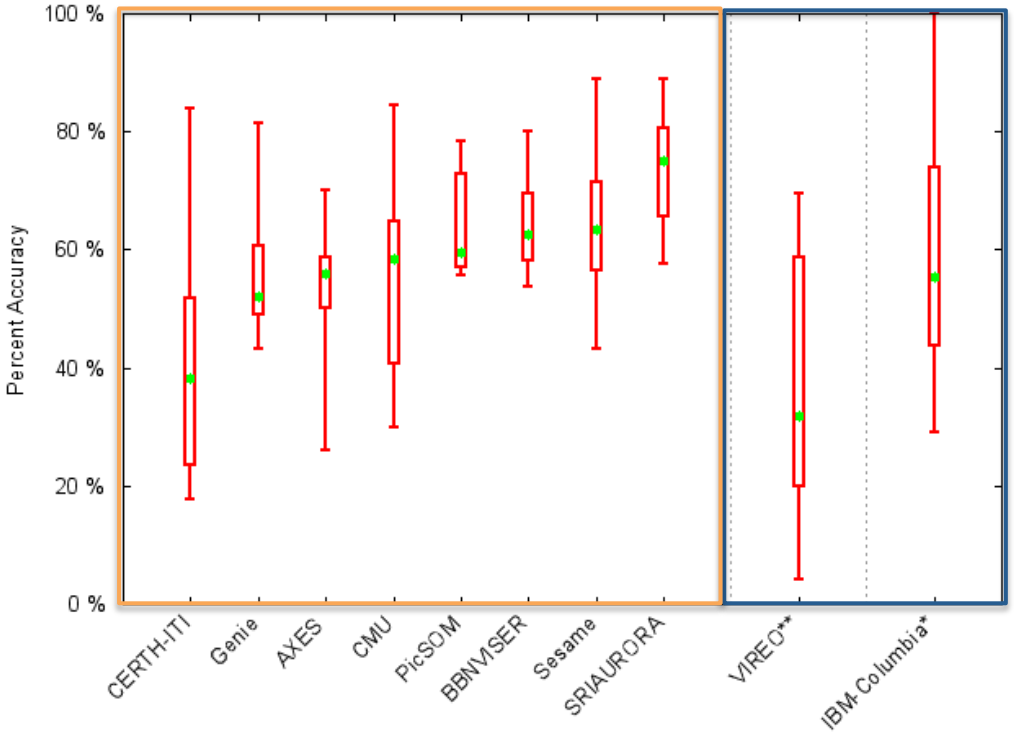
- Recountings were selected for event:clip pairs where ideally a recounting would be present for each system.
  - 6 clips per event were chosen, 4 true positives, and 2 false positives.
  - Prioritized event:clip pairs that were a part of last years progress set.
  - Up to 60 recountings per system could be present, on average ~55 recountings.
    - For 2 systems, different sets of recountings had to be selected.
  - Order of presentation was counterbalanced across teams and events.
- 10 Judges: each was assigned ~3 recountings per team, per event.
- Using the NIST provided MER evaluation application judges are instructed to study the event kit text then assess the recounting by ..
  - *For each observation ..*
    - Reading the observation text.
    - Viewing/hearing *all* of the snippets.
    - Grading the observation text, as a description of the snippets.
  - On the basis of the recounting, deciding whether or not the clip contains an instance of the event

The Metrics

- Accuracy
- Percent Recounting Review Time (PRRT)
- Precision of the observation text (Observation Text Score)

Metrics:Accuracy

By Team

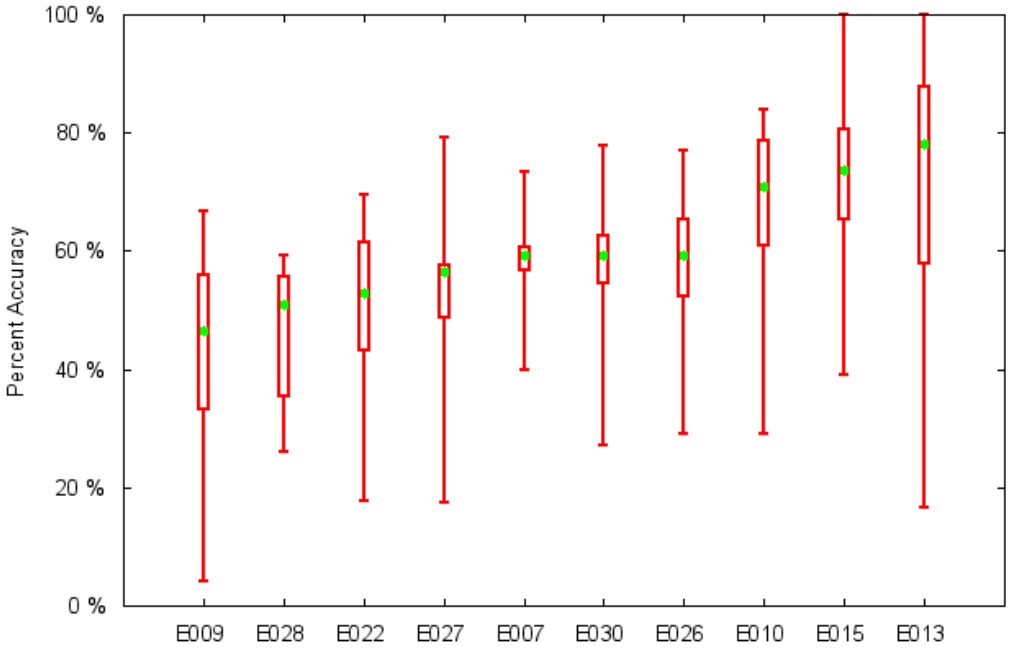


* -> Results generated from a different set of test clips
** -> Results generated from a different set of test clips

- Question posed to judges:
  - Based on this recounting, does the clip contain an instance of the event?
    - The clip *contains an instance* of the event
    - The clip *does not* contain an instance of the event
    - I do not know because the *recounting* does not allow me to tell whether the clip contains an instance of the event
    - I do not know because the *event kit* does not allow me to tell whether the clip contains an instance of the event
- The degree to which the judges' assessments agree with the MED ground truth.
  - (Number of correctly labeled clips / number of clips judged as one of the first three options specified above)

Metrics:Accuracy

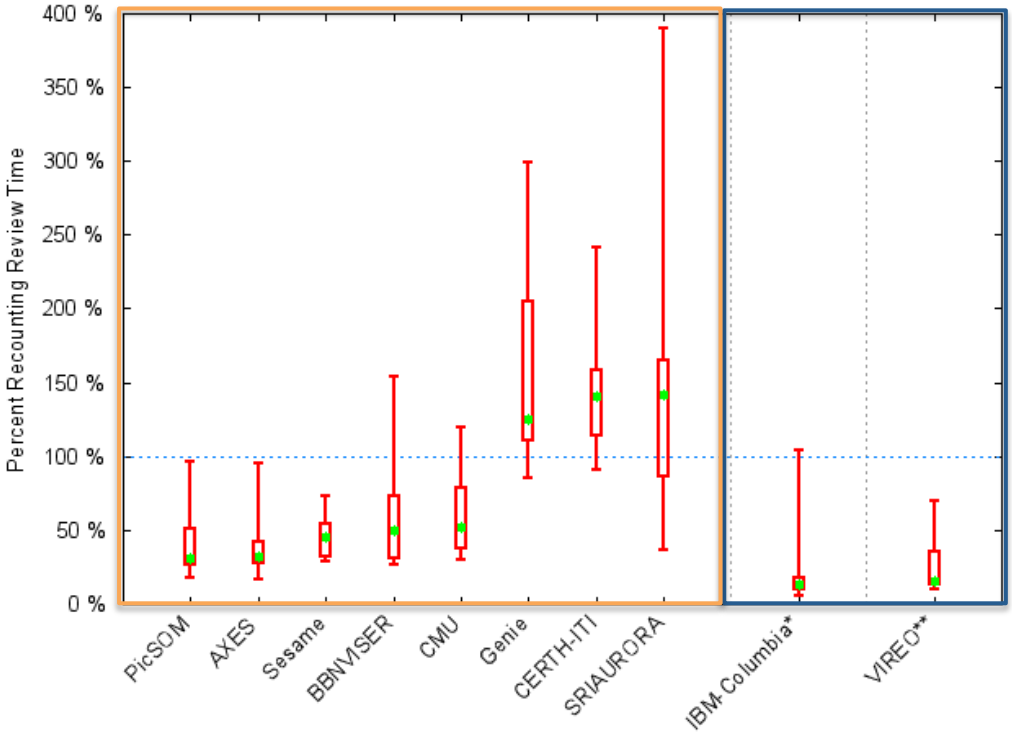
By Event



- E009 Getting a vehicle unstuck
- E028 Town hall meeting
- E022 Cleaning an appliance
- E027 Rock climbing
- E007 Changing a vehicle tire
- E030 Working on a metal crafts project
- E026 Renovating a home
- E010 Grooming an animal
- E015 Working on a sewing project
- E013 Parkour

Metrics:PRRT

By Team

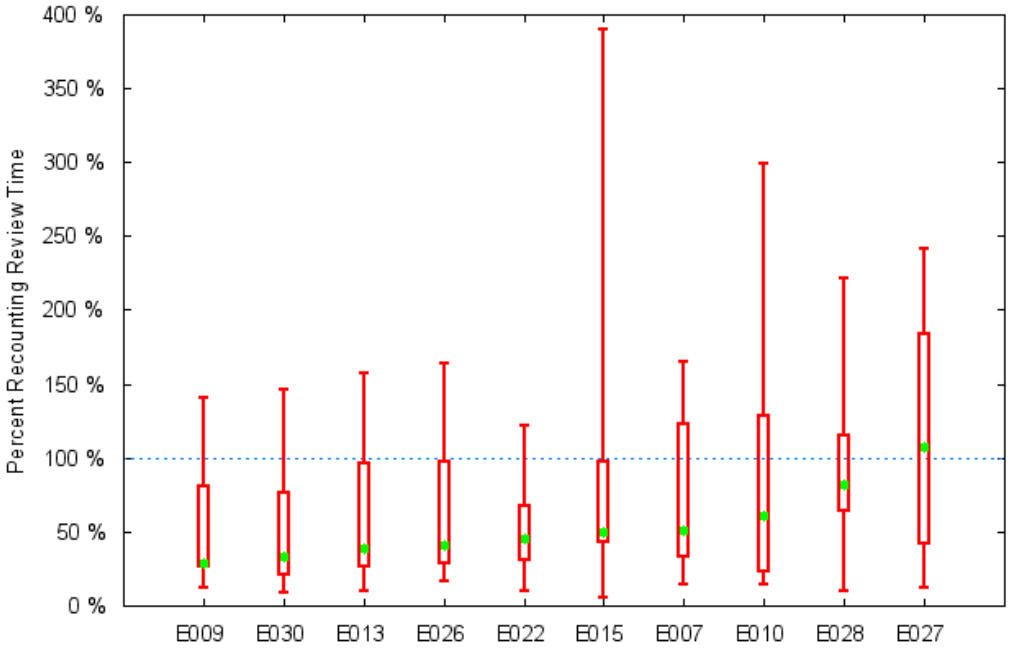


- The percentage of clip time the judges took to judge the recounting.
  - (Total judgment time / Total duration of clips to be assessed)

* -> Results generated from a different set of test clips
** -> Results generated from a different set of test clips

Metrics:PRRT

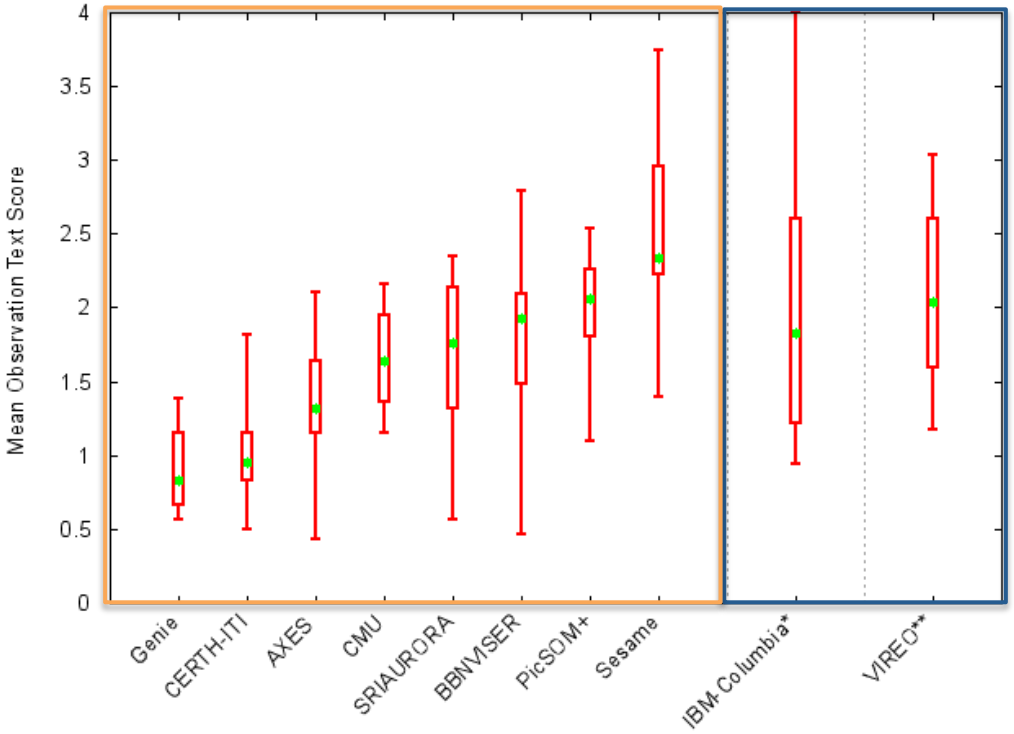
By Event



- E009 Getting a vehicle unstuck
- E030 Working on a metal crafts project
- E013 Parkour
- E026 Renovating a home
- E022 Cleaning an appliance
- E015 Working on a sewing project
- E007 Changing a vehicle tire
- E010 Grooming an animal
- E028 Town hall meeting
- E027 Rock climbing

Metrics:Observation Text Score

By Team

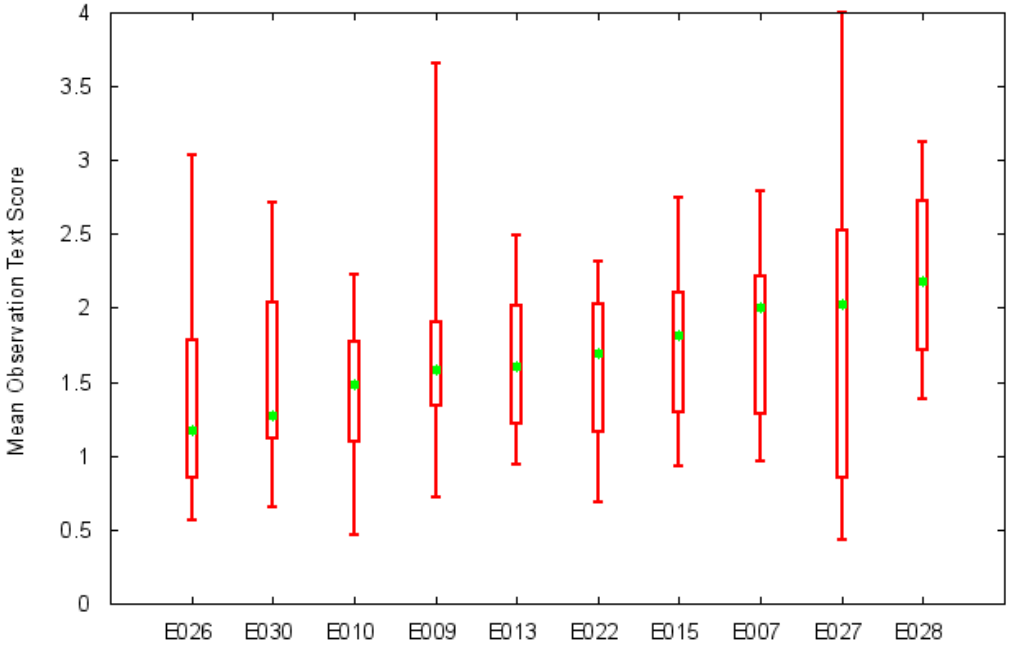


- Question posed to judges:
  - How well does the text of this observation describe the snippet(s)?
    - A: Excellent (4 points)
    - B: Good (3 points)
    - C: Fair (2 points)
    - D: Poor (1 point)
    - F: Fails (0 points)
- The mean of the judges' scores given to the observations.

* -> Results generated from a different set of test clips
** -> Results generated from a different set of test clips
+ -> Observation text was always "visual content matches examples"

Metrics:Observation Text Score

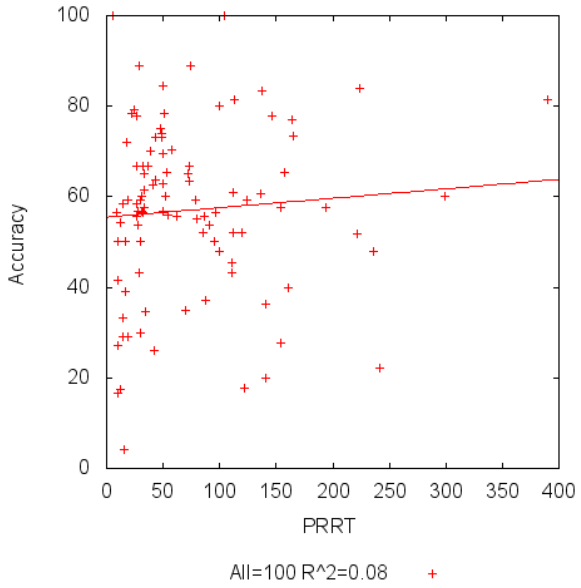
By Event



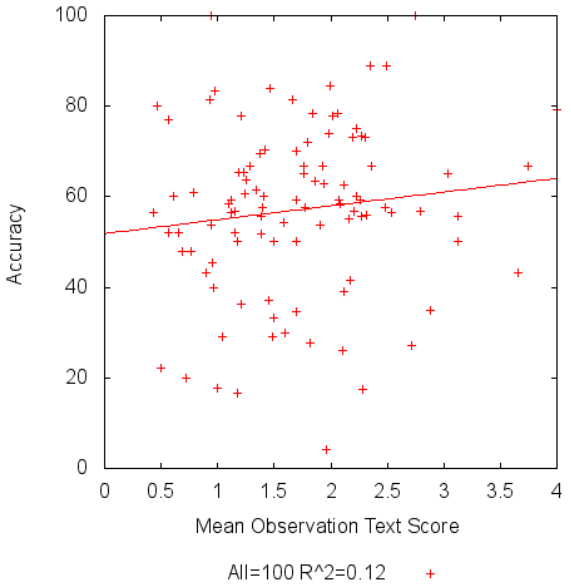
- E026 Renovating a home
- E030 Working on a metal crafts project
- E010 Grooming an animal
- E009 Getting a vehicle unstuck
- E013 Parkour
- E022 Cleaning an appliance
- E015 Working on a sewing project
- E007 Changing a vehicle tire
- E027 Rock climbing
- E028 Town hall meeting

Metric Analysis

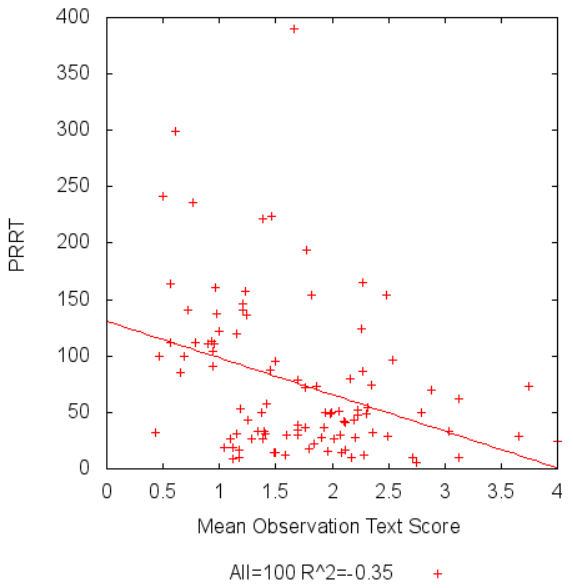
Accuracy vs. PRRT



Accuracy vs. Mean Observation Text Score



PRRT vs. Mean Observation Text Score



Judge Feedback

- Some judges found the wording of the questions to be confusing and unclear
- Some found that the task was tedious and difficult at times

Conclusion

- Purpose of MER systems:
  - Recount the evidence in a multimedia clip for a specific event
- MER metrics fairly independent of each other
- Judging process and training could be refined